Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR

Sandra Wachter*

Oxford Internet Institute, University of Oxford and The Alan Turing Institute, British Library, London, United Kingdom

ABSTRACT

In the Internet of Things (IoT), identification and access control technologies provide essential infrastructure to link data between a user's devices with unique identities, and provide seamless and linked up services. At the same time, profiling methods based on linked records can reveal unexpected details about users' identity and private life, which can conflict with privacy rights and lead to economic, social, and other forms of discriminatory treatment. A balance must be struck between identification and access control required for the IoT to function and user rights to privacy and identity. Striking this balance is not an easy task because of weaknesses in cybersecurity and anonymisation techniques. The EU General Data Protection Regulation (GDPR), set to come into force in May 2018, may provide essential guidance to achieve a fair balance between the interests of IoT providers and users. Through a review of academic and policy literature, this paper maps the inherit tension between privacy and identifiability in the IoT. It focuses on four challenges: (1) profiling, inference, and discrimination; (2) control and context-sensitive sharing of identity; (3) consent and uncertainty; and (4) honesty, trust, and transparency. The paper will then examine the extent to which several standards defined in the GDPR will provide meaningful protection for privacy and control over identity for users of IoT. The paper concludes that in order to minimise the privacy impact of the conflicts between data protection principles and identification in the IoT, GDPR standards urgently require further specification and implementation into the design and deployment of IoT technologies.

Keywords: Data protection, Digital Ethics, Identity, Identification, Internet of things, Privacy, Profiling, Discrimination, GDPR, Review

* Dr. Sandra Wachter, Oxford Internet Institute, University of Oxford, 1 St. Giles, Oxford, OX1 3JS United Kingdom
Email address: sandra.wachter@oii.ox.ac.uk Tel:+44(0)7478340679

1 Introduction

Usage of the 'Internet of Things' (IoT)[1] is rapidly growing. The European Union expects major investments in areas such as smart homes, personal wellness and wearables, smart energy, smart cities, and smart mobility.[2] IoT applications are emerging across myriad sectors, for example in healthcare,[3] energy consumption and utility monitoring,[4] transportation and traffic control,[5] logistics,[6] production and supply chain management,[7] agriculture,[8] public space and environmental monitoring,[9] social interactions,[10] personalised shopping and commerce,[11] domestic automation,[12] and others. These IoT devices constantly collect vast amounts of personal data such as location data and health data (e.g. FitBit) in order to function properly or to optimise and customise their services.

The IoT is defined by connections and linked services, tailored to the specific requirements of users. Objects and services must be connected to one another and share data about a specific user to provide networked services that are informed by more than the user's direct interaction with a particular node. Without repeated and consistent identification of users, linked up, seamless services would not be possible.

---

[1] Defining the 'Internet of Things' is not straightforward. As argued by Whitemore et al. based on a 2015 literature survey, a core concept of the IoT is that "everyday objects can be equipped with identifying, sensing, networking and processing capabilities that will allow them to communicate with one another and with other devices and services over the Internet to achieve some useful objective" (Andrew Whitmore, Anurag Agarwal and Li Da Xu, 'The Internet of Things--A Survey of Topics and Trends' (2015) 17 Information Systems Frontiers; New York 261, 261). While seemingly any Internet-connected object can be treated as part of the IoT, narrower definitions are also available. In logistics and supply chain management, the Internet of Things can refer simply to 'objects' embedded with RFID tags, allowing for unique identification and monitoring of object movement and consumption (Whitmore, Agarwal and Da Xu; Rolf H Weber, 'Internet of Things ? New Security and Privacy Challenges' (2010) 26 Computer Law & Security Review 23.). The term is also often used as a synonym for ubiquitous computing or ambient intelligent, referring to "smart devices, sensors, human beings, and any other object that is aware of its context and is able to communicate with other entities" (Farzad Khodadadi, Amir Vahid Dastjerdi and Rajkumar Buyya, 'Internet of Things: An Overview' [2017] arXiv preprint arXiv:1703.06409 <https://arxiv.org/abs/1703.06409> accessed 30 June 2017.). In other words, the IoT can refer to a network of sensing objects that monitor and record aspects of their environment and the behaviours of users within it. Alongside well-established RFID tags, wireless sensor networks and Bluetooth-enabled devices have emerged as IoT sensors.

[2] European Commission, 'Commission Staff Working Document: Advancing the Internet of Things in Europe' (European Commission 2016) SWD(2016) 110 final 31 <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016SC0110&from=EN> accessed 8 July 2017.

[3] Khodadadi, Dastjerdi and Buyya (n 1); F Gonçalves and others, 'Security Architecture for Mobile E-Health Applications in Medication Control', *2013 21st International Conference on Software, Telecommunications and Computer Networks - (SoftCOM 2013)* (2013); Cisco, 'Securing the Internet of Things: A Proposed Framework' (2016) <http://www.cisco.com/c/en/us/about/security-center/secure-iot-proposed-framework.html> accessed 6 July 2017.

[4] S Sicari and others, 'Security, Privacy and Trust in Internet of Things: The Road Ahead' (2015) 76 Computer Networks 146; Khodadadi, Dastjerdi and Buyya (n 1).

[5] Sicari and others (n 4); Khodadadi, Dastjerdi and Buyya (n 1).

[6] Sicari and others (n 4); C Yuqiang, G Jianlan and H Xuanzi, 'The Research of Internet of Things' Supporting Technologies Which Face the Logistics Industry', *2010 International Conference on Computational Intelligence and Security* (2010).

[7] Sicari and others (n 4); L Weiss Ferreira Chaves and C Decker, 'A Survey on Organic Smart Labels for the Internet-of-Things', *2010 Seventh International Conference on Networked Sensing Systems (INSS)* (2010).

[8] Khodadadi, Dastjerdi and Buyya (n 1).

[9] Sicari and others (n 4); Khodadadi, Dastjerdi and Buyya (n 1).

[10] Khodadadi, Dastjerdi and Buyya (n 1).

[11] Sicari and others (n 4).

[12] Khodadadi, Dastjerdi and Buyya (n 1).

However, the pursuit of identification and personalisation of users poses a risk to privacy. Data controllers can draw inferences from these data.[13] Users can easily perceive this insight as invasive, unexpected, and unwelcome. Discriminatory treatment can also result from inferential analytics and linkage of disparate records,[14] motivating limitations on user profiling.[15] The impossibility of anonymising data[16] and weak cybersecurity standards (often owing to the limited computational power of identifying technologies such as WiFi or RFID)[17] can exacerbate privacy risks.

Together, these risks make free and well-informed consent challenging in the IoT. Privacy policies often fail to communicate clearly the risks of data processing and linkage of user records (which requires consistent user identification).[18] The EUs General Data Protection Regulation (GDPR) might improve the situation. The regulation will come into force in May 2018, and accounts for many of these risks. The GDPR creates governing principles of data processing (Articles 5 and 25) and establishes new data protection standards relevant for IoT devices. New harmonised standards relating to informed consent, notification duties, privacy by design and privacy by default, data protection impact assessment, algorithmic transparency, automated decision-making, and profiling will apply across Europe.

These standards will be undermined by the tendency of IoT devices and services to collect, share, and store large and varied types of personal data, to operate seamlessly and covertly, and to personalise functions based on prior behaviour.

This paper analyses the inherit tension between privacy and identifiability in IoT by reviewing prior discussion in academic and policy discourse. Four topics are identified which describe the nature and effects of privacy challenges arising from identity management in the IoT: (1) profiling, inference, and discrimination; (2) control and context-sensitive sharing of identity; (3) consent and uncertainty; and (4) honesty, trust, and transparency. Key issues and potential solutions to balance privacy and identifiability are analysed in the context of new requirements and protections introduced by the GDPR. The analysis suggests that new approaches to transparency and user awareness will be crucial to balance privacy and identifiability, while accounting for potential discrimination, weaknesses in security and anonymisation, and poorly informed consent. Rather than promising that privacy can always be guaranteed in the IoT, transparency, awareness, and honesty are needed about the possible risks (e.g. via notifications, or access rights). Without open communication of the risks inherent to the IoT, informed consent and informational self-determination will be hindered.

The paper is structured as follows: section 2 describes the role of identification technologies in the IoT to provide linked up, personalised services. Section 3 then examines the tension between user privacy and linkage enabled by IoT identification technologies. Section 4 then reviews how academic

[13] Sarah Johanna Eskens, 'Profiling the European Citizen in the Internet of Things: How Will the General Data Protection Regulation Apply to This Form of Personal Data Processing, and How Should It?' (Social Science Research Network 2016) SSRN Scholarly Paper ID 2752010 <https://papers.ssrn.com/abstract=2752010> accessed 8 July 2017.
[14] Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104 California Law Review.
[15] Sandra Wachter, 'Privacy: Primus Inter Pares — Privacy as a Precondition for Self-Development, Personal Fulfilment and the Free Enjoyment of Fundamental Human Rights' (Social Science Research Network 2017) SSRN Scholarly Paper ID 2903514 <https://papers.ssrn.com/abstract=2903514> accessed 19 February 2017.
[16] Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006> accessed 28 June 2017.
[17] R Roman, P Najera and J Lopez, 'Securing the Internet of Things' (2011) 44 Computer 51.
[18] Omer Tene and Jules Polonetsky, 'Big Data for All: Privacy and User Control in the Age of Analytics' [2013] Nw.J. Tech. & Intell. Prop. <http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/nwteintp11&section=20> accessed 2 October 2014.

literature and policy dealt with the challenges of user privacy and critically analyses the extent to which the GDPR addresses these issues. Section 5 concludes that future directions for research, design and business practices must seek to strike a balance between privacy and identifiability. In order to minimise the privacy impact of the conflicts between data protection principles and identification in the IoT, several GDPR standards urgently require further specification and implementation into the design and deployment of IoT technologies.

2 Identification technologies

The IoT is built on the principle that connected nodes should be able to communicate and share data with one another when authorised. Objects should exchange data in real-time to provide a linked up service for the user,[19] with appropriate limitations and control points to protect user privacy preferences. Devices and services operate on different local and global networks, which are governed by different technical and international legislative standards, and developed by different manufacturers.[20] Linkage between disparate devices and services requires trusted communication between networked objects and users.

Identification technologies are a crucial component of trusted communication in the IoT. To be trustworthy, communication should occur only between intended nodes in the network. Consistent and unique identification of objects and users is required to ensure devices communicate only with their intended targets, and are not exposed to third party attacks, data leakage, or communication with unintended devices and users. 'Identity management' describes identification technologies that assign, manage, or verify such unique identities to establish and maintain trust in communication between IoT objects and users. Identity management enables resource (i.e. tracking objects) and service discovery (i.e. communication and access to data).[21]

Various systems using centralised and distributed architectures can manage identities.[22] There is no universally adopted identity management standard yet,[23] but certain standards have gained widespread usage.[24] IoT developers often prefer centralised identity and access management systems due to their convenience for users. However, they can face challenges relating to scalability and cross-border governance harmonisation, and pose privacy risks to users because they allow greater exchange and linkage of potentially sensitive personal data between device or service providers.

[19] N Karimian, PA Wortman and F Tehranipoor, 'Evolving Authentication Design Considerations for the Internet of Biometric Things (IoBT)', *2016 International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)* (2016).

[20] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Transparent, Explainable, and Accountable AI for Robotics' (2017) 2 Science Robotics.

[21] Khodadadi, Dastjerdi and Buyya (n 1).

[22] Michal Trnka and Tomas Cerny, 'Authentication and Authorization Rules Sharing for Internet of Things' (2017) 2017 Software Networking 35, for example, describe a "central identity store" for identity management that uses OAuth 2.0 and OpenID Connect tokens in all device communication. Single sign-on systems (e.g. Keycloak, OpenID, Shibboleth) are also widely used. Devices are granted a unique identity in the database, which can be linked to a separate user identity or profile.

[23] Roman, Najera and Lopez (n 17). A fundamental challenge of the IoT is to design technical and semantic mechanisms that enable interoperability and shared meaning in communications between devices and services. These challenges can be re-imagined as an issue of international governance, given the IoT's global reach.

[24] A prominent standard is the Object Naming Service (ONS), based on the Domain Name Service (DNS) model, which uses an Electronic Product Code (EPC) to retrieve information about an object. See: Yanjiong Wang and Qiaoyan Wen, 'A Privacy Enhanced Dns Scheme for the Internet of Things'.

As a basis for trusted communication, identity management is often coupled with access control[25] to authenticate nodes in a network for operations, transmission, and access to data.[26] Together, these functions enable identity to serve as a basis for privacy-enhancing authentication schemes.[27] IoT developers must set access permissions not only for users, but also for 'things' seeking to access or process a user's data on the user's or a third party's behalf. These access permissions must respect user's privacy preferences.[28] Identity management combined with role-based access control, for example, enables identity verification coupled with authorisation of users' actions requests or devices according to their system-assigned role,[29] ensuring that only actions authorised to a specific role (e.g. collecting, transmitting, or processing data) can be taken in a session.[30] Administrators or users can define these permissions, offering different approaches to protect user's subjective privacy preferences.

Numerous authentication systems (e.g. OpenID, PKI, SAML, Ucode) have emerged that support this pairing of identity management and access control.[31] Biometric authentication in particular has grown in recent years,[32] based on inputs such as activity, behavioural modelling via accelerometer data,[33] electrocardiogram (ECG) signals (AppleWatch, for example, uses interruptions in heartbeat as a trigger for authentication), fingerprint (e.g. Identilock to prevent unauthorised users from firing guns), vein recognition, and iris recognition.[34] While potentially more secure than traditional username/password combinations,[35] these data sources introduce new risks of invasive inferences due to the inherent sensitivity of health data.[36] Similar concerns arise for authentication based on behavioural modelling, which can trigger re-authentication if it detects abnormal behavioural or usage patterns. At the same time, granular, potentially invasive modelling of user behaviours is required for this type of authentication.

Consistent management of identity and access permissions is complicated because IoT users can enter and leave networks repeatedly, and move across networks, and national boundaries. Users must

[25] Access control systems can rely on centralised or distributed identity stores, and utilise various forms of authentication including anonymous protocols that hide a user's identity from data controllers. See: Almudena Alcaide and others, 'Anonymous Authentication for Privacy-Preserving IoT Target-Driven Applications' (2013) 37 Computers & Security 111.
[26] Trnka and Cerny (n 22); Sicari and others (n 4); Cisco (n 3).
[27] Roman, Najera and Lopez (n 17).
[28] Sicari and others (n 4). These permissions can be managed through various systems, based on a "hierarchical model, reputation mechanisms, approaches derived from social networking, fuzzy techniques, mechanisms based on nodes past behaviour or on routing strategies."
[29] Trnka and Cerny (n 22).
[30] Roman, Najera and Lopez (n 17); Ravi S Sandhuy, 'Role-Based Access Control' <https://pdfs.semanticscholar.org/5108/3705c268568b0e691cfa443f6cd03ff43416.pdf> accessed 6 July 2017.
[31] Sicari and others (n 4). Approaches to authentication include traditional username and password combinations, IP address, device tagging or fingerprinting (i.e. a unique identifier modelled on usage data and device characteristics), and biometrics. Radio Frequency Identification (RFID) tags, for example, enable objects to be uniquely identified or 'tagged' using radio frequency signals. RFID can be used to authenticate a device for data access, for instance allowing a medical practitioner to retrieve a patient's medical records by scanning an attached RFID tag. See: Gonçalves and others (n 3).
[32] Cisco (n 3).
[33] S Batool, NA Saqib and MA Khan, 'Internet of Things Data Analytics for User Authentication and Activity Recognition', *2017 Second International Conference on Fog and Mobile Edge Computing (FMEC)* (2017).
[34] Karimian, Wortman and Tehranipoor (n 19).
[35] A Alterman, '"A Piece of Yourself": Ethical Issues in Biometric Identification' (2003) 5 Ethics and Information Technology 139.
[36] Brent Daniel Mittelstadt and Luciano Floridi, 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts' (2016) 22 Science and Engineering Ethics 303; D Cantore, 'On Biometrics and Profiling: A Challenge for Privacy and Democracy?' (2011) 2 International Journal of Technoethics 84.

connect not only to central nodes or identity stores and databases, but also communicate with one another based on commonly understood access policies.[37] The diversity of possible systems for identity management and access control poses a challenge for harmonisation of privacy and security standards across the IoT. Globally accepted identity management protocols, "certification authorities", and "a common-accepted trust negotiation language" do not yet exist.[38] While diversification can be beneficial for cybersecurity purposes (by avoiding a single potential point of failure or attack), it can simultaneously increase the burden on users seeking to verify that the standards governing a particular device or service are fit-for-purpose, and thus do not pose undue risks to their privacy.

3 User identity and linked services

To understand how identification technologies in the IoT pose a risk to user privacy, it is helpful to appeal to Floridi's concept of informational identity[39] and its segmentation into virtual identities used to interact with the IoT.[40] Informational identity is the sum of the information that exists about the individual. Such information can be of many types, including legal representations of identity (e.g. driver's license, passport or national identity card, contracts), digital representations of identity (e.g. legal representations in addition to online user accounts, social networks, preferences and attributes inferred from usage or behavioural data), and any other information describing the person. Information and communication technologies create new flows of information describing users, which impact, on how the user self-identifies, and how others perceive and understand them.[41]

Privacy is inherently connected to the integrity of the information constituting one's identity, insofar as privacy facilitates the individual to control information about herself, and thereby manage external transformation of this information. The right to privacy is, in the words of Warren and Brandeis, "the right to one's personality."[42] Such external management of identity, seen for example in the creation of knowledge about an individual through inferential profiling, constitutes an attack on the integrity of the individual's identity.[43] Floridi's approach acknowledges the inherent invasiveness of user profiling and inferential analytics detached from the user's awareness or oversight of these processes. Such external transformations of identity create difficulties for a narrative approach to personal identity, which treats identity as a self- or socio-constructed series of connected and coherent pieces of information.[44] In other words, profiling creates and modifies information about the user, which becomes part of the individual's identity and shapes her future treatment within IoT systems, and more broadly within the infosphere.[45] When a user is unaware that her IoT devices and services

[37] Connected cars must, for example, communicate with one another on the road to ensure safe operation. See: Khodadadi, Dastjerdi and Buyya (n 1). See also ibid., who argue that federation can reduce the complexity of authentication introduced by growth in the scale of a network. Rather than authenticating for each new service, users verify their identity once with a trusted domain which then authenticates on the user's behalf with other users and domains. Single-sign on (SSO) systems are a common form of federated authentication. In the IoT, SSO allows for devices to be linked to specific users on a semi-permanent basis. A semi-permanent trusted relationship can be established in this way, allowing the device to operate on the user's behalf in collaboration with other devices and services.

[38] Sicari and others (n 4) 154.

[39] Luciano Floridi, 'The Informational Nature of Personal Identity' (2011) 21 Minds and Machines 549.

[40] Amardeo C Sarma and João Girão, 'Identities in the Future Internet of Things' (2009) 49 Wireless Personal Communications 353.

[41] Luciano Floridi, 'The Informational Nature of Personal Identity' (2011) 21 Minds and Machines 549.

[42] Samuel D Warren and Louis D Brandeis, 'The Right to Privacy' (1890) 4 Harvard Law Review 193, 33.

[43] Floridi, 'The Informational Nature of Personal Identity' (n 39).

[44] Marya Schechtman, *The Constitution of Selves* (Cornell university press 2007).

[45] The infosphere, or informational environment, "is a context constituted by the whole system of information objects, including all agents and patients, messages, their attributes and mutual relations." Luciano Floridi, 'On the Intrinsic Value of Information Objects and the Infosphere' (2002) 4 Ethics and information technology 287,

generates information about her, she lacks the ability to incorporate this information into her self-constructed identity (or narrative), and to view herself as others view her (with the inaccessible information).[46] Even if accessible, personal information that ICTs externally construct may only be available in aggregate, rather than as a distinct event in a series of connected and coherent events within a third party's narrative. The risks of identity management in the IoT to user control of personality can thus be better appreciated when considering its impact on both the self-constructed and externally-constructed informational identity of users.

Sarma and Girão describe a similar approach. According to them, digital identities (for a user, device, etc.) can be viewed as a sum of constructs that describe entities. These constructs include characteristics such as "age, identifiers, and in general concepts, propositions and claims about an identity."[47] It follows that digital identities that IoT construct and mediate are not merely a sum of offline identifiers such as age, name, address, but rather a sum of all descriptive characteristics about an entity. IoT devices and services generate data that can be transformed into information describing an entity, and constituting that entity's identity.[48] The user's identity is constructed and transformed through the addition of constructs or information describing her.[49]

Digital identities in the context of the IoT are best understood as a type of profile, made up of all information describing the user that is accessible to a decision-maker, based on observations or prior knowledge (e.g. age, location), or inferences about the user (e.g. behaviours, preferences, predicted future actions). Digital identity both uniquely 'singles out' users (i.e. for authentication) and contains information (e.g. inferences) about them. The user may remain unaware of the full extent and nature of her identity, as the information constituting this identity can be externally constructed (i.e. through inferential analytics).

Therefore, both the individual (i.e. how the individual views and understands herself), and external entities (e.g. a social network provider making inferences based on usage data) construct and manage an individual's informational identity. Identification technologies help organise this external management of identity. As external construction implies, the user may lack control or oversight of the content of their identity, how it changes over time and shapes their experiences when the identity is known to other users.[50] Further, users are often unable to assess the validity and quality of inferences made about them. As a result, the externally constructed identity often remains unclear to the user.

A user's digital identity can be segmented for various purposes. Only portions of the user's identity or profile are typically available to specific entities.[51] These segments of a user's overall identity or profile can be referred to as 'virtual identities'. As Sarma and Girão explain:

> *"Virtual identity denotes an entity in a specific role or usage context, not the entity in its entirety…a virtual identity is a sub-set of the digital information about a user. A user may*

289. For a fuller discussion, see: Luciano Floridi, 'Information Ethics: On the Philosophical Foundation of Computer Ethics' (1999) 1 Ethics and Information Technology 33; Luciano Floridi, *The Ethics of Information* (Oxford University Press 2013).

[46] Floridi, 'The Informational Nature of Personal Identity' (n 41).

[47] Sarma and Girão (n 40) 354.

[48] Floridi, 'The Informational Nature of Personal Identity' (n 39).

[49] For extensive discussion of the tension between profiling and identity see: Mireille Hildebrandt, 'Profiling and the Identity of the European Citizen' [2008] Profiling the European citizen 303.

[50] Floridi, 'The Informational Nature of Personal Identity' (n 39).

[51] Amardeo Sarma and others, 'Virtual Identity Framework for Telecom Infrastructures' (2008) 45 Wireless Personal Communications 521.

have more than one virtual identity to represent the different personas and aspects of its service usage… A virtual identity may contain data relevant to many services and networks, including subscriptions, identifiers, service preferences, etc. While its logical understanding is that the information may be available from any point in the architecture, the information itself can be distributed and limited to access by only some entities."[52]

A virtual identity can be associated simultaneously with multiple devices via authentication, and passed downstream to devices and services. As the identity is passed across and between networks, it constrains actions by other nodes on behalf of the user, and her preferences (e.g. how much information to share with other users in the network). Sarma and Girão refer to this as the user's 'digital shadow'.[53] By sharing a virtual identity with a device, the user is able to influence and constrain its operation. However, the virtual identity only needs to fulfil a context-specific function, which does not normally require disclosure of a user's full identity or profile. The user's context-specific virtual identity thus allows the user to interact with multiple devices seamlessly under a single profile.

The virtual identity concept was originally proposed as a privacy-enhancing mechanism for identity management in the IoT.[54] Well-curated virtual identities, used as a digital shadow, can reveal only the information that a user deems necessary for delivery of a particular service, or operation of a particular device, without disclosing the user's complete legal or digital identity. The virtual identity "contains information about his attributes and the objects' sessions and interactions with the architecture…digital shadows only implicitly indicate their owner's identity."[55]

A desire to help users retain control over their data, as well as interactions on their behalf with other nodes in IoT networks, is implicit in the concept of a 'virtual identity'. However, there is an implicit tension between user's information privacy (or control of personal data)[56] and IoT, as the latter relies on linked up services. To identify users, IoT controllers and third parties will need to link data about specific users from disparate sources, with or without the consent of the user. Identity management systems grant only authorised access and communication between trusted and intended users, while also providing the necessary infrastructure for potentially invasive linkage of virtual identities and profiling. As the IoT requires linkage between devices and services for personalisation, a fundamental tension exists between linkage of IoT nodes, and privacy.

4 Privacy and IoT identification technologies

The previous section has sketched the fundamental tension between privacy and linkage via identity management in the IoT. It remains open how this tension manifests in practical challenges for design and regulation of the IoT. To map these challenges, a review of relevant academic and policy literature was conducted[57] using four databases (Web of Science, IEEE, Google Scholar, Google).

[52] Sarma and Girão (n 40) 354.
[53] ibid.
[54] ibid; Sarma and others (n 51).
[55] Roman, Najera and Lopez (n 17).
[56] J van Hoof and others, 'Ambient Intelligence, Ethics and Privacy' (2007) 6 Gerontechnology <http://gerontechnology.info/index.php/journal/article/view/704> accessed 21 September 2012; Luciano Floridi, 'Four Challenges for a Theory of Informational Privacy' (2006) 8 Ethics and Information Technology 109; Bart W Schermer, 'The Limits of Privacy in Automated Profiling and Data Mining' (2011) 27 Computer Law & Security Review 45.
[57] The review addresses privacy and trust as normative concepts. Prior surveys are available detailing technical and legal aspects of the IoT relevant to privacy and trust. Atzori et al. have reviewed open issues relating to standardization, addressing, and networking (Luigi Atzori, Antonio Iera and Giacomo Morabito, 'The Internet

Specifically, prior work discussing (1) the IoT, (2) identity management or identification of devices and users, and (3) privacy was reviewed. In addition to systematic and consistent database searching, additional sources were identified by hand and through cross-referencing of the sample returned via the database queries. In total, 60 sources were reviewed in full. Papers were read by the author, and key passages were highlighted and grouped into types of challenges facing the IoT. The review identifies major challenges and developments around privacy and trust in identity management and profiling. The literature is benchmarked against the new standards in the GDPR to assess if the new data protection framework is able to resolve the inherent tension between privacy and identifiability in the IoT.

A thematic structure was developed inspired by recent academic work and European policy priorities, including the GDPR. In an influential paper on regulating the IoT, Scott R. Peppet identified four major areas of concern going forward: "discrimination, privacy, security, and consent." The paper explains that American regulation and policies is insufficient guard against these potential risks.[58] European policy has identified similar areas of concern: "minimisation, purpose limitation, data retention/deletion, automated decision taking/profiling (including discrimination), and security requirements."[59] IoT devices collect vast amounts of data often without the awareness of the data subjects. Users often have very little control over their data, and lack sufficient knowledge to give free and informed consent. Pervasive data collection, AI-based analyses, and the combination of datasets pose serious risks for privacy as data controllers can draw invasive inferences about the user.[60]

Recognising these risks, four types of challenges are used to structure the review, which roughly follow Peppet's taxonomy.[61] Based on themes that emerged from the reviewed literature, there are at least four senses in which identification technologies pose a risk to user privacy: (1) by enabling linkage of user identities and records generated from IoT devices, which can lead to potentially invasive profiling, inferences, and discrimination; (2) by revealing sensitive information to other IoT users that the user would otherwise prefer to keep confidential, and inhibiting user's control over such disclosures; (3) by generating information or inferences about the user, which could not have been predicted when setting access policies, or consented to at the point of adoption; and (4) by limiting

---

of Things: A Survey' (2010) 54 Computer Networks 2787). Miorandi et al. have reviewed open challenges for security requirements relating to data confidentiality, privacy, and trust (Daniele Miorandi and others, 'Internet of Things: Vision, Applications and Research Challenges' (2012) 10 Ad Hoc Networks 1497). Weber has reviewed legislative requirements for privacy in the IoT (Weber (n 1).). Finally, Roman et al. and Sicari et al. have reviewed security and privacy in the IoT, focusing in particular on new cybersecurity risks (Rodrigo Roman, Jianying Zhou and Javier Lopez, 'On the Features and Challenges of Security and Privacy in Distributed Internet of Things' (2013) 57 Computer Networks 2266; Sicari and others (n 4).).

[58] Scott R Peppet, 'Regulating the Internet of Things: First Steps toward Managing Discrimination, Privacy, Security and Consent' (2014) 93 Tex. L. Rev. 85.

[59] W Kuan Hon, Christopher Millard and Jatinder Singh, 'Twenty Legal Considerations for Clouds of Things' 23 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2716966> accessed 8 July 2017; European Commission, 'Report on the Public Consultation on IoT Governance' (2013) <http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=1746> accessed 8 July 2017; Monica Salgado, 'Internet of Things Revisited' (2014) 15 Privacy and Data Protection 12; Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) 0829/14/EN WP216 29 <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf> accessed 8 July 2017.

[60] Hon, Millard and Singh (n 59) 23; European Commission, 'Radio Frequency Identification (RFID) in Europe: Steps towards a Policy Framework' (2011) <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:l24120a> accessed 8 July 2017.

[61] Although Peppet's work primarily discussed challenges of the IoT in an American context, his analysis of the IoT itself applies regardless of location. The thematic review conducted here is inspired by his analysis and proposed taxonomy, but adapts and applies it to a European policy context.

user oversight and transparency in management of identity and profiling, which can facilitate breaches of privacy and undermine trust among IoT users, objects, and device or service providers. Each type of challenge is reviewed in detail in the remainder of this section.

While understanding the tension between privacy and identity management in the IoT is useful in itself, solutions are also required. To this end, the paper considers the extent to which each of these challenges may be resolved by design choices, business practices and regulation. Particular attention is paid to the GDPR, which will overhaul data protection law in Europe in 2018, and will impose new standards on IoT providers and data controllers to protect user privacy. Specifically, this paper considers the relevant GDPR standards for transparency (Article 5), data storage, access, rectification, and deletion (Articles 5, 15-17), informed consent (Article 7), notification duties (Articles 13-14 and 33), automated decision-making and profiling (Articles 21-22), privacy by design and privacy by default (Article 25), cybersecurity (Articles 33-34), and data protection impact assessment (Article 35-36).

4.1 Profiling, inference, and discrimination

There are at least three possible ways of monitoring and profiling that offer grounds for discrimination in IoT systems: (a) data collection that leads to inferences about the person (e.g. internet browsing behaviour); (b) profiling at large through linking IoT datasets (sometimes called 'sensor fusion'); and (c) profiling that occurs when data are shared with third parties that combine data with other datasets (e.g. employers, insurers).

According to Roman et al., users can have "access to an unprecedented number of personalized services, all of which would generate considerable data, and the environment itself would be able to acquire information about users automatically."[62] Unpredictable, invasive profiling and inferential analytics can result from data sharing, in particular when multiple IoT devices provide data that are linked to a single user (virtual) identity.

Identification technologies enable precisely this sort of linkage. By linking multiple devices and the data they produce to a single user identity, the usage of a device or service can be personalised, based upon past behaviours and preferences, and inferences drawn from these data.[63] Privacy risks of linkage between datasets become particularly acute when central authentication systems (e.g. SSO) or identity stores have access to data that authenticated devices generate. While potentially offering a 'better' user experience, linkage and personalisation across multiple IoT devices and services nonetheless pose risks to user privacy. Data controllers can draw inferences about the user unrelated to the intended operation of the devices and services.[64] Device identification can be used to link together a user's behaviours, even if each of the datasets is individually handled responsibly and appropriately de-identified. Algorithms can be used to update continuously user profiles to predict their behaviour and match their preferences.[65]

Part of the challenge of controlling profiling is the uncertain value of data that sensors generate. Peppet describes this as 'sensor fusion'.[66] In addition to the stream of data collected, the possible inferences drawn can be broader if combined with other data categories from the IoT. For example, Fitbit opens insight into the user's health status (e.g. heart rate), as well as into user's movement (e.g.

[62] Roman, Najera and Lopez (n 17).
[63] Schermer (n 56).
[64] ibid.
[65] Tal Zarsky, 'Transparent Predictions' (2013) 2013 University of Illinois Law Review <http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2324240> accessed 17 June 2016.
[66] Peppet (n 58) 93.

geolocation data, steps taken per day).[67] Combining these two data streams could lead to further privacy invasive inferences.[68] Similarly, Peppet argues that "existing smartphone sensors can be used to infer a user's mood; stress levels; personality type; bipolar disorder; demographics (e.g., gender, marital status, job status, age); smoking habits; overall well-being; progression of Parkinson's disease; sleep patterns; happiness; levels of exercise; and types of physical activity or movement'.[69] Eskens makes a similar point about the 'Nest' brand thermostat, which "collects data such as current temperature, humidity, ambient light, and whether something in the room is moving." Based on these data, which are collected to adjust temperature automatically, inferences can be made about the presence and specific location of occupants in a home, their current state (e.g. asleep or awake), and other aspects of home activity.[70] These examples illustrate that smart devices need to collect data and make inferences (e.g. is the person at home, does the temperature need to be adjusted) in order to work properly,[71] but this can simultaneously and inadvertently lead to intrusions into the private lives of users.

While some inferences and profiling drawn from IoT can be benign – for example when data are used to provide a more personalised user experience – they can also lead to unfair discrimination (e.g. economic or gender based).[72] The potential for discrimination holds true even when using non-sensitive data categories, from which sensitive information can still be inferred.[73] Third parties with access to IoT data linked to an identified target can use these data for purposes with which the user would not agree if asked. FitBit data could, for example, be relevant to prospective employers, who could make inferences about "impulsivity and the inability to delay gratification-both of which might be inferred from one's exercise habits-correlate with alcohol and drug abuse, disordered eating behaviour, cigarette smoking, higher credit-card debt, and lower credit scores. Lack of sleep-which a Fitbit tracks-has been linked to poor psychological well-being, health problems, poor cognitive performance, and negative emotions such as anger, depression, sadness, and fear."[74]

Employers are not the only third parties potentially interested in this data. Boerman et al. explain that data controllers increasingly use the IoT for "monitoring people's online behaviour and using the information collected to show people individually targeted advertisements."[75] The authors warn that a lack of awareness of such methods can be unfair to users, hence an increased level of transparency is required. Turow similarly maintains that opaque profiling and automated decision-making in advertisements can pose a threat to diversity.[76] Even seemingly neutral data (e.g. postcodes) can lead

[67] ibid 121; Mahmoud Elkhodr, Seyed Shahrestani and Hon Cheung, 'A Review of Mobile Location Privacy in the Internet of Things', *ICT and Knowledge Engineering (ICT & Knowledge Engineering), 2012 10th International Conference on* (IEEE 2012).
[68] Peppet (n 58).
[69] ibid 115–116.
[70] Eskens (n 13) 19.
[71] ibid 20.
[72] Peppet (n 58); Latanya Sweeney, 'Discrimination in Online Ad Delivery' (2013) 11 Queue 10:10; Tene and Polonetsky (n 18); Katja de Vries, 'Identity, Profiling Algorithms and a World of Ambient Intelligence' (2010) 12 Ethics and Information Technology 71.
[73] Andrea Romei and Salvatore Ruggieri, 'A Multidisciplinary Survey on Discrimination Analysis' (2014) 29 The Knowledge Engineering Review 582.
[74] Peppet (n 58) 119.
[75] Sophie C Boerman, Sanne Kruikemeier and Frederik J Zuiderveen Borgesius, 'Online Behavioral Advertising: A Literature Review and Research Agenda' (2017) 0 Journal of Advertising 1.
[76] Joseph Turow, *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth* (Yale University Press 2012).

to inference and discrimination[77] based on ethnicity, gender, or sexual preference[78] especially when data sets are linked.[79]

Weaknesses of anonymisation to prevent profiling and resulting discrimination cause further problems. According to Gudgel, "there is special concern that if data is not anonymized then it could potentially be used to track specific individuals, linked to information in other databases, and possibly used to predict future behaviour."[80] Tracking data of the type that many IoT devices generate is notorious to open to re-identification and reverse engineering of identity.[81] Following the assumption that data cannot be permanently anonymised without destroying their analytical value,[82] non-technical methods may be necessary to prevent profiling and discrimination in the IoT. One potential solution is to treat all IoT generated data that refer to a user as personal data under data protection law,[83] as it will always be possible in principle to link the data back to a person.[84] This approach would ensure that the user would be able to exercise her rights granted under data protection law over all information that IoT devices create and manage. This concept would not prevent profiling as a result; but rather, extend the scope of existing user rights against privacy risks to cover all data, including inferences and profiles.

These problems will be exacerbated by the proliferation of machine learning in the IoT.[85] Machine learning will lead to even less predictable inferences, while the complexity and opaqueness of machine learning algorithms can inadvertently hide discriminatory treatment from users.[86] Systems operating as 'black boxes', for which the inputs, internal logic, and outputs may be unavailable or incomprehensible to individual users do not facilitate systematic observation, identification of harmful effects, or investigation of their causes.[87] Machine learning can inadvertently and unknowingly reinforce existing biases and prejudices as a result.[88]

European legislators have addressed the risks of profiling and discrimination, albeit often lacking detailed recommendations.[89] European regulators have raised "major concerns in declarations on

[77] Barocas and Selbst (n 14).
[78] Schermer (n 56); Kevin Macnish, 'Unblinking Eyes: The Ethics of Automating Surveillance' (2012) 14 Ethics and Information Technology 151.
[79] Salvatore Ruggieri, Dino Pedreschi and Franco Turini, 'Data Mining for Discrimination Discovery' (2010) 4 ACM Transactions on Knowledge Discovery from Data (TKDD) 9; Tal Zarsky, 'The Trouble with Algorithmic Decisions An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making' (2016) 41 Science, Technology & Human Values 118.
[80] John Gudgel, 'Objects of Concern? Risks, Rewards and Regulation in the "Internet of Things"' (Social Science Research Network 2014) SSRN Scholarly Paper ID 2430780 12 <https://papers.ssrn.com/abstract=2430780> accessed 8 July 2017.
[81] Peppet (n 58).
[82] Ohm (n 16); Kyle Ebersold and Richard Glass, 'THE INTERNET OF THINGS: A CAUSE FOR ETHICAL CONCERN.' (2016) 17 Issues in Information Systems <http://www.iacis.org/iis/2016/4_iis_2016_145-151.pdf> accessed 12 June 2017; Gianmarco Baldini and others, 'Ethical Design in the Internet of Things' [2016] Science and Engineering Ethics 1.
[83] Eskens (n 13) 21; Ohm (n 16); Peppet (n 58).
[84] Tene and Polonetsky (n 18); Ohm (n 16).
[85] Peppet (n 58) 122.
[86] Jenna Burrell, 'How the Machine "Thinks:" Understanding Opacity in Machine Learning Algorithms' [2016] Big Data & Society.
[87] ibid; Brent Mittelstadt, 'Auditing for Transparency in Content Personalization Systems' (2016) 10 International Journal of Communication 12.
[88] Tene and Polonetsky (n 18).
[89] For a discussion on different AI governance efforts see Corinne Cath and others, 'Artificial Intelligence and the "Good Society": The US, EU, and UK Approach' [2017] Science and Engineering Ethics 1.

profiling,"[90] acknowledging that profiling offers grounds for discrimination, especially when datasets are combined. The European Commission has called for the creation of a set of guiding principles to govern IoT regulation, urging that "always being connected to the things around us has the potential to lead to more surveillance or more profiling by public authorities and private entities."[91] Similarly, the European Data Protection Supervisor has voiced concerns that RFID tags used in IoT systems could lead to profiling by linking users to specific devices and usage records.[92]

Similar concerns are reflected in the GDPR, especially in Article 21 (Right to object) and Article 22 (Automated individual decision-making, including profiling). Article 21 introduces the right of data subjects to object to data processing, including profiling, at any time. If the purpose of data processing is direct marketing, the data subject will have an absolute right to object. In all other cases data processing must stop, unless the data controller can demonstrate compelling legitimate interests that override the interests of the data subjects. Unfortunately, the framework does not offer a definition of compelling interests of data controllers,[93] leaving both data controllers and data subjects in an uncertain situation. Apart from this uncertainty, the technical feasibility of ceasing data collection is also challenging. How data controllers can handle objections beyond ceasing all service provision remains unclear. As a result, users concerned about their privacy or IoT-facilitated profiling may be left with a binary 'take it or leave it' choice.

Article 22 introduces additional safeguards against automated decision-making, including profiling, but only when data processing is solely automated and has legal or similar significant effects. The scope of applicability is thus likely to be very limited, at least while these terms ('solely automated', 'legal or similarly significant effects') remain undefined in practice.[94] In cases where such decision-making and profiling are necessary for entering or performing a contract between data subject and the data controller, or based on explicit consent (Article 22(2)(a) and (c)), data subjects are granted rights to obtain human intervention on the part of the controller, to express a point of view and to contest the decision (Article 22(3)). If automated decision-making, including profiling, has significant effects on a data subject, individuals will possess a legal remedy if they disagree with the outcome. Finally, at first sight Art 11 GDPR appears to be helpful. This provision echoes the idea of only identifying data subjects for as long as necessary. However, as stated above, discrimination is also possible through extraneous additional, non-personal or anonymous data. In those cases, data protection law either does not apply or offers insufficient protection. With a broader and well-defined scope of applicability, these rights would offer a very promising approach for data subjects to maintain some control over how the data is used to personalise services and future opportunities.

4.2 Control and context-sensitive sharing of identity

Users are often powerless to prevent potentially discriminatory profiling due to a lack of control over disclosures of personal data and identity in the IoT.[95] Segmentation of a user's overall identity can involve the creation of multiple virtual identities that are used to connect to specific networks,

[90] Hon, Millard and Singh (n 59) 19.
[91] European Commission, 'Commission Staff Working Document: Advancing the Internet of Things in Europe' (n 2) 27.
[92] European Commission, 'Radio Frequency Identification (RFID) in Europe: Steps towards a Policy Framework' (n 60).
[93] Martini, 'DS-GVO Art. 21 Widerspruchsrecht' in Paal and Pauly (eds), *Datenschutz-Grundverordnung* (1st edn, beck-online 2017) Rn. 37-40.
[94] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' [2017] International Data Privacy Law <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469>.
[95] Mittelstadt and Floridi (n 36).

devices, or services.[96] Mechanisms to support targeted disclosure of identity may thus facilitate protection of privacy. However, this benefit can only be claimed when meaningful control over disclosures is feasible; granting disclosure control to users who do not understand the potential risks and benefits of revealing different aspects of their identity can actually expose users to greater privacy risks. Generic or provider-defined disclosure policies may thus be preferable in some cases.

As it stands, users often lack any ability to define such context-sensitive constraints on identity disclosure. Devices acting on the user's behalf, using their virtual identity, often similarly lack comparable context-sensitive constraints on identity and personal data disclosure.[97] Automatic service discovery protocols can broadcast sensitive information, including device hostname and the device owner's identity, which the user may prefer to keep secret. Monitoring devices challenge privacy because they generate "traceable signatures of the location and behaviour" of individuals.[98] Users thus often lack granular control over the initiation of sessions with IoT networks.[99] Inadvertently broadcasted identities enable linking users to devices, to virtual identities, and subsequent profiling.[100] Wu et al. describe a possible solution in a private service discovery protocol based on encrypted identity certificate exchange to protect the privacy of users and service providers. The protocol makes services discoverable only to authorised clients, which grants users some control over identity disclosure.[101]

Multi-user and multi-controller models of ownership of objects also challenge control of identity and personal information disclosure in the IoT. A single-user, single-device model can no longer be assumed.[102] Devices can have multiple identities across networks using different identification standards and access controls. Similarly, multiple users can use the same device, meaning a device identity will not be connected to a single user identity. The inverse is also true.[103] Usage and control of the operation of a device will not always centre on an individual user. All members of a household or fitness club, for example, can use a smart scale, not just one individual. Cases involving a single controller but multiple users are also common; sensing devices embedded in public spaces or work environments can monitor multiple individuals, but very often a single entity controls them.

Models for segmenting control have implications for privacy and confidentiality of the data of individuals and groups of users.[104] It can be expected, for example, that a multi-user device will grant users access only to their individual data in the interest of privacy and confidentiality.[105] To limit access to data records and profiles only to intended users, user identity must be confirmed before data are recorded. To ensure data confidentiality and integrity, it is thus a common function of IoT devices to authenticate users prior to collecting data or providing access to previously collected data.[106]

---

[96] Sarma and others (n 51).
[97] Roman, Najera and Lopez (n 20) provide the example of an undercover police car in a connected vehicle network needing to temporarily hide its policing function from other vehicles.
[98] Cisco (n 3).
[99] Sarma and Girão (n 40) 356.
[100] David J Wu and others, 'Privacy, Discovery, and Authentication for the Internet of Things' [2016] arXiv:1604.06959 [cs] <http://arxiv.org/abs/1604.06959> accessed 30 June 2017.
[101] ibid.
[102] Sarma and Girão (n 40).
[103] ibid.
[104] Brent Mittelstadt, 'From Individual to Group Privacy in Big Data Analytics' (2017) 30 Philosophy & Technology 475; Lee A Bygrave, *Data Protection Law: Approaching Its Rationale, Logic and Limits* (Kluwer Law Intl 2002); Alessandro Mantelero, 'Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection' (2016) 32 Computer Law & Security Review 238.
[105] Cisco (n 3).
[106] Sicari and others (n 4).

Similarly, for purposes of authorising and controlling communication and interoperability, IoT devices are normally uniquely identifiable (e.g. an individual connected light bulb), or identifiable as a group (e.g. all smart thermostats in a building).[107] For multi-user, multi-controller, and multi-device networks, fair segmentation of control between these entities remains a challenge.

The challenges for individuals to have meaningful control over personal data are echoed in the GDPR. The rights based approach of the GDPR aims to equip users with the necessary tools to manage usage of their data. Article 15 (Right of access), Article 16 (Right to rectification), and Article 17 (Right to erasure) are notable in this regard. The right of access (Article 15) grants data subjects the right to request at any time information about the types of personal data that a data controller processes, and to receive a copy of the processed data (Article 15(3)). However, rights and freedoms of others (e.g. privacy, trade secrets, Art 15(4), Recital 63) can trump the latter. The right of access follows the principle of transparency (Article 5). The right to rectification (Article 16) grants users to rectify inaccurate data and complete incomplete data, which is an essential tool for informational self-determination. Finally, to mitigate undesired data usage, data subjects will – under certain circumstances (Article 17(1)) – be able to request erasure of their data, for example, when the data are no longer essential for purposes for which they were collected. These rights foster the implementations of the GDPR's guiding principle of transparency (Article 5).

At first glance, these rights appear promising. Nonetheless, as is the case with 'legitimate interests' of data controllers, clarity is lacking on how the interests of data controllers and users will be balanced. Users have a clear interest in disclosure control to prevent privacy violations and discriminatory treatment, which must be balanced with the rights and freedoms of others (including data controllers). Unfortunately, these rights also do not require data controllers to indicate precisely which data were (most) relevant to a particular decision or effect of identification or profiling. As a result, individuals may have to comb through thousands of entries to identify potentially inaccurate, incomplete, or misleading data.[108] This situation creates an additional barrier to meaningful control of informational identity. The actual impact of these rights on user control over identity disclosure will primarily be established through future European jurisprudence, which will clarify how these interests must be balanced in practice.

4.3 Consent and uncertainty

The uncertain value of data that IoT devices generate challenge user-centric access control (including 'virtual identities' and 'digital shadows') is and service as a privacy preserving mechanism. Privacy- and trust-enhancing identity management and access control systems frequently use a user-centric definition of privacy and trust.[109] Accordingly, a system is considered privacy- and trust-enhancing if it grants the user oversight and choice over the way in which their IoT devices communicate and take actions on their behalf, and thus how their IoT-generated data are shared with IoT devices, other users, and data controllers. Enforcement of privacy preferences prior to communication between IoT devices can protect context-sensitive and subjective expectations of privacy.[110] As Roman, Najera,

[107] Cisco (n 3).
[108] Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' [2017] arXiv preprint arXiv:1711.00399.
[109] Tene and Polonetsky (n 18); Michael Friedewald and others, 'Privacy, Identity and Security in Ambient Intelligence: A Scenario Analysis' (2007) 24 Telematics and Informatics 15; F Massacci, VH Nguyen and A Saidane, 'No Purpose, No Data: Goal-Oriented Access Control Forambient Assisted Living' (2009) <http://www.scopus.com/inward/record.url?eid=2-s2.0-74049137043&partnerID=40&md5=98e37b22f5acce7ffaf8f6bdac154020>.
[110] Baldini and others (n 82); Martin Henze and others, 'A Comprehensive Approach to Privacy in the Cloud-Based Internet of Things' (2016) 56 Future Generation Computer Systems 701. See also: Helen Nissenbaum,

and Lopez argue that effective permissions for communication, data sharing, and processing must in some way reflect the preferences and interests of the user.[111]

Several approaches to user-centric access control were noted in the reviewed literature. Privacy by design seeks to provide users with tools to oversee and control how they generate and share data from IoT devices and services.[112] Dynamic consent tools already exist to allow granular control over data access.[113] Another example is a 'privacy coach': an RFID reader embedded in a mobile phone that scans an object and downloads its privacy policy, then compares it to the user's privacy preferences and provides a recommendation about whether the device meets the user's requirements, and thus whether it should be used.[114]

As argued above, IoT identification technologies enable linkage of user profiles, and unpredictable inferences.[115] If a user's network-specific virtual identity is linked back to the user's other virtual, digital or legal identities, fragmentation of identity fails as a privacy preserving mechanism. Users may be unaware of the scope and granularity of their data, and its potential value and inferences that can be drawn from it, as well as the extent to which their data are accessible to third parties outside the context or purpose for which it was created.[116] The uncertain risks accompanying identification technologies, coupled with the conflicting need for users to make an informed choice when setting access permissions (or, at a higher level, choosing to use an IoT device or service in the first place), undermines the actual protection that user-centric identity management and access controls can offer. Communicating this uncertainty to users remains an outstanding challenge for IoT controllers seeking informed consent.[117]

If the uncertainty of inferential analytics prevents users from making an informed choice when adopting and using an IoT system (including setting subjective privacy-preserving access permissions), informed consent may be infeasible. The capacity of data subjects to freely consent will further be challenged if they cannot fully understand the scope due to the complexity of privacy policies.[118] Tene and Polenetsky[119] argue that the term 'privacy policy' is misleading, since users may believe that data controllers want to protect their privacy and not share their data, when in fact these policies serve as liability disclaimer.

'Privacy as Contextual Integrity' (Social Science Research Network 2004) SSRN Scholarly Paper ID 534622 <http://papers.ssrn.com/abstract=534622> accessed 12 March 2013.

[111] SI Ahamed, N Talukder and Md M Haque, 'Privacy Challenges in Context-Sensitive Access Control for Pervasive Computing Environment' (2007) <http://www.scopus.com/inward/record.url?eid=2-s2.0-50249180369&partnerID=40&md5=da4e8dd7956060a59a64ebf6dac97ecf>.

[112] Roman, Najera and Lopez (n 17).

[113] ibid; Jane Kaye and others, 'Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks' (2015) 23 European Journal of Human Genetics 141.

[114] Gerben Broenink and others, 'The Privacy Coach: Supporting Customer Privacy in the Internet of Things' [2010] arXiv:1001.4459 [cs] <http://arxiv.org/abs/1001.4459> accessed 7 July 2017.

[115] Mittelstadt and Floridi (n 36); Vries (n 72); Eskens (n 13).

[116] danah boyd and Kate Crawford, 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon' (2012) 15 Information Communication & Society 662; Kerstin Denecke and others, 'Ethical Issues of Social Media Usage in Healthcare' (2015) 10 IMIA Yearbook of Medical Informatics 137; Brent Mittelstadt, 'Ethics of the Health-Related Internet of Things: A Narrative Review' (2017) 19 Ethics and Information Technology 157.

[117] Joshua Fairfield and Hannah Shtein, 'Big Data, Big Problems: Emerging Issues in the Ethics of Data Science and Journalism' (2014) 29 Journal of Mass Media Ethics 38; Peppet (n 58).

[118] Peppet (n 58) 148.

[119] Tene and Polonetsky (n 18).

Access permissions can be set in many ways. Users can express high-level preferences, which are automatically translated into granular, device- or user-specific permissions. Trust-based access control schemes[120] can, for example, calculate a trust score and assign permissions based on the score. For example, in the fuzzy trust based access control scheme design by Mahalle et al., a trust score is generated based on experience, knowledge, and recommendations from other IoT nodes. These factors are linked to a user's assessment of trust, and thus her choice to grant communication with the node. More explicit embedding of user preferences in setting access permissions is also possible; role-based access control schemes can, for example, give users explicit granular control over permissions.

For access control schemes that enhance user privacy and trust, the specifics of how permissions are set are irrelevant for our current purposes. Rather, what matters is the user-centric conception of privacy and trust that is implicit or explicit in such schemes. Schemes enhance privacy or trust when permissions and authorizations reflect a user's subjective preferences and interests.[121] It follows that the actual protection that such schemes offer depends upon the quality of a user's choice in setting high- or low-level permission preferences. Similar to informed consent, user-centric models are only effective to that extent that users are able to express their interests related to privacy and trust, are informed of the potential risks of permitting communication or data sharing between IoT nodes, and are made aware of downstream risks resulting from these actions.

In relation to privacy protection and informed consent, the GDPR will set new standards across Europe. Article 25 creates a general duty for data controllers to implement privacy by default and privacy by design mechanisms. This could also help to resolve the uncertainty of privacy invasive analytics and thus offer a better basis for informed consent. If the user is assured that privacy will be protected by default, the user can make an informed choice as the potential privacy consequences become foreseeable. The specific measures required will depend on the circumstances. Article 25 states that data controllers need to take "state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons" into account when personal data are processed. General principles (reflected in Article 5) must also govern technical and organisational measures, such as data minimisation, storage and purpose limitations. Each measure faces implementation barriers. The 'data maximalism' reflected in the IoT's built-in reliance on vast data collection and sharing for personalised services suggests minimalism will be unlikely, or at least require significant trade-offs in terms of functionality. Data controllers may need to consider designing IoT services that, by default, require as little data as possible to operate. It may be necessary to 'start over' at the design phase, rather than just adopting current business practices, as perpetual storage and limitless re-purposing of collected data will directly conflict with the GDPR's guiding principles. However, a privacy by design approach, even if challenging, could help to increase user trust if data controllers are seen to protect their privacy by default.

Nonetheless, even if privacy by default and privacy by design mechanisms are implemented, the problem of uncertainty remains for informed consent mechanisms. Unforeseen inferences can be routinely drawn,[122] even if privacy is otherwise protected (e.g. via encryption or pseudonymisation). To inform about the potential risks of data collection, the GDPR has created higher standards relating to informed consent (Article 7) and notification duties (Article 13-14). In addition to the need to

[120] PN Mahalle and others, 'A Fuzzy Approach to Trust Based Access Control in Internet of Things', *Wireless VITAE 2013* (2013).
[121] Ahamed, Talukder and Haque (n 111); Broenink and others (n 114).
[122] Burrell (n 86).

communicate the possible risks of data processing in "in an intelligible and easily accessible form, using clear and plain language" (Article 7(2)), Article 7 also aims to prevent excessive data collection. As Article 7(4) states, the assessment of whether consent was freely giving will also depend on whether "personal data that is not necessary for the performance of that contract" was part of consent. Excessive data collection can thus render consent invalid. A similar constraint is found in Article 13(2)(e), which explains that prior to data collection, the controller must state "whether the provision of personal data is a statutory or contractual requirement, or a requirement necessary to enter into a contract, as well as whether the data subject is obliged to provide the personal data and of the possible consequences of failure to provide such data." This provision creates a legal requirement to minimise excessive data collection despite existing tendencies towards data maximalism. The concerns against excessive data collection are also reflected in the Art 29 Working Party's guidelines on consent[123] that promote granular and customised consent to readjust the power asymmetry between data controllers and data subjects, and to move away from the current "take it or leave it" approach. Finally, long and complicated privacy notices have proven to be ineffective, as users often do not read them. The GDPR might improve the situation as Art 12(7) explains that "The information to be provided to data subjects pursuant to Articles 13 and 14 may be provided in combination with standardised icons in order to give in an easily visible, intelligible and clearly legible manner a meaningful overview of the intended processing." This makes it clear that the new framework prefers short text with icons over long privacy statements. A similar view is reflected in Art 12(8), which states that the European Commission is tasked to develop standardised icons. Art 13 and 14 thus aim to make users immediately aware of the intended data collection without expecting users to read elaborate privacy notices.

4.4 Honesty, trust, and transparency

Identity management aims to establish and maintain trust between objects and users in the IoT.[124] Trust relationships for sharing data can be defined at a device-to-device, device-to-user, and device-to-controller level, with permissions attached to the identity of specific devices, users, and controllers. Trust between objects refers to authentication prior to communication and data access. Prevention of unauthorised objects and users from accessing a system can enhance confidentiality, and thus increase user trust.[125]

In contrast, trust between users and objects addresses user perceptions and impressions of control. Users often believe that trust is a prerequisite for systems to be privacy enhancing.[126] According to Yan et al., user "trust is a complicated concept with regard to the confidence, belief, and expectation

[123] Art 29 Working Party, 'Guidelines on Consent under Regulation 2016/679, 17/EN WP25; Adopted on 28 November 2017' <https://iapp.org/media/pdf/resource_center/wp29_consent-12-12-17.pdf> accessed 4 February 2018.
[124] C Perera and others, 'Context Aware Computing for The Internet of Things: A Survey' (2014) 16 IEEE Communications Surveys Tutorials 414.
[125] Sicari and others (n 4).
[126] S Chakraborty, H Choi and MB Srivastava, 'Demystifying Privacy in Sensory Data: A QoI Based Approach' (2011) <http://www.scopus.com/inward/record.url?eid=2-s2.0-79958057222&partnerID=40&md5=e907b7a413b53870811659ac9d8b0682>; U Rashid, H Schmidtke and N Woo, 'Managing Disclosure of Personal Health Information in Smart Home Healthcare' (2007) <http://www.scopus.com/inward/record.url?eid=2-s2.0-38149103634&partnerID=40&md5=6e8a9a957ac835d8d2ce6949be2b3a71>; Kai Wang and others, 'Pervasive and Trustworthy Healthcare', *Advanced Information Networking and Applications - Workshops, 2008. AINAW 2008. 22nd International Conference on* (2008); Sicari and others (n 4).

on the reliability, integrity, security, dependability, ability, and other characters of an entity."[127] To achieve trust in an IoT system, developers must demonstrate how they plan to mitigate the significant legal and ethical risks associated with their devices. Developers often currently fail to demonstrate compliance with data protection rules, for example through accreditation programmes.[128]

Processing data in a way that complies with user's rights and expectations can enhance user trust.[129] Perceptions of being watched,[130] helplessness, and external control can impact on adoption of IoT.[131] To combat these perceptions, Roman et al. argue that users must be helped to retain control and oversight with "tools that accurately describe all their interactions so that they can form an accurate mental map of their virtual surroundings."[132] Weber and Weber also suggest that "individuals should be able to disconnect from their networked environment at any time."[133] Scholars have proposed comparable opt-out options for profiling practices.[134]

Similarly, Weber argues that transparency plays an essential role in increasing users' trust and suggests several steps to increase transparency:

> *"Provide information about the intended collection, storage and/or data processing; provide an overview of what personal data have been disclosed to what data controller under which policies; provide online access to the personal data and how they have been processed; and provide counter profiling capabilities helping the user to anticipate how their data match relevant group profiles, which may affect future opportunities or risks."*[135]

The European Data Protection Supervisor has similarly argued that transparency is key to ensure that users are aware of data collection and processing.[136] The European Commission has taken a related position, arguing that increased trust can be achieved through compliance with data protection and cybersecurity law, and implementation of transparency tools, privacy-enhancing technologies, and safeguards against profiling and monitoring.[137] These actions are essential since key challenges facing the IoT are "security, liability, privacy and data protection."[138]

---

[127] Zheng Yan, Peng Zhang and Athanasios V Vasilakos, 'A Survey on Trust Management for Internet of Things' (2014) 42 Journal of Network and Computer Applications 120.
[128] Kit Huckvale and others, 'Unaddressed Privacy Risks in Accredited Health and Wellness Apps: A Cross-Sectional Systematic Assessment' (2015) 13 BMC Medicine 214.
[129] Sicari and others (n 4).
[130] S Welsh and others, 'Big Brother Is Watching You - The Ethical Implications of Electronic Surveillance Measures in the Elderly with Dementia and in Adults with Learning Difficulties' (2003) 7 Aging and Mental Health 372.
[131] Roman, Najera and Lopez (n 17).
[132] ibid.
[133] Rolf H Weber and Romana Weber, *Internet of Things: Legal Perspectives*, vol 49 (Springer Science & Business Media 2010) 39.
[134] Alan Rubel and Kyle ML Jones, 'Student Privacy in Learning Analytics: An Information Ethics Perspective' (Social Science Research Network 2014) SSRN Scholarly Paper ID 2533704 <http://papers.ssrn.com/abstract=2533704> accessed 22 July 2015; Mireille Hildebrandt, 'Who Needs Stories If You Can Get the Data? ISPs in the Era of Big Number Crunching' (2011) 24 Philosophy & Technology 371; Mittelstadt, 'Ethics of the Health-Related Internet of Things' (n 116).
[135] Rolf H Weber, 'Internet of Things: Privacy Issues Revisited' (2015) 31 Computer Law & Security Review 618, 625; similarly, Tene and Polonetsky (n 18).
[136] European Commission, 'Radio Frequency Identification (RFID) in Europe: Steps towards a Policy Framework' (n 60).
[137] European Commission, 'Commission Staff Working Document: Advancing the Internet of Things in Europe' (n 2) 28–29.
[138] ibid 30.

The GDPR has several provisions that enhance trust and transparency. Data controllers will need to carry out a data protection impact assessment (DPIA) (Article 35) if they, for example, use new technologies, such as profiling methods. DPIA's can be useful to increase user trust, especially if they are published publicly, as recommended by the Article 29 Working Party.[139] Ideally, a DPIA should evaluate the associated risks of data processing and propose solutions to mitigate identified risks.

Further, Article 13(2)(f) and 14(2)(g) aim to inform data subjects about the "the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject." These provisions are intended to empower data subjects to evaluate the risks when algorithmic methods are used prior to data processing (Art12(7)), e.g. via privacy statements on webpages.[140] This is also reflected in the Art 29 Working Party's guidelines on transparency[141] as they state "Recital 39 stipulates, amongst other things, that data subjects should be "made aware of the risks, rules, safeguards and rights in relation to the processing of personal data and how to exercise their rights in relation to such processing". For all of these reasons, WP29's position is that data controllers should provide the information to data subjects well in advance of the stipulated time limits." The guidelines suggest that data subjects should be made aware of the risks of profiling as soon as possible to enable them to exercise their rights. However, it is uncertain whether notification alone can achieve this goal. The functionality of machine learning and the unpredictability of inferences drawn from large data sets can be difficult to understand for lay people as well as experts. Given that Article 12(7) states that the information in Article 13-14 can be provided with icons, the intended level of sophistication for the information provided appears to be low. It is thus questionable whether the notification duties will provide data subject with meaningful understanding of the risks of machine learning. The same holds true for the DPIA, as possible risks of machine learning might not be anticipated or anticipatable.

The GDPR also relates trust to cybersecurity standards. Having a secure system meeting at least recognised minimal standards can increase user trust. To this end, Article 5 defines the general principle of integrity and confidentiality. Likewise, Articles 33-34 describe how data controllers must react to data breaches. In case of a breach, controllers must inform the Supervisory Authority within 72 hours.[142] In cases of breaches posing a high-risk data controllers have to inform the data subjects. Each of these provisions can contribute to transparency and a trusting relationship between users and IoT device and service providers. However, as notification is only required in the case of 'high risk' breaches, how this threshold is defined generally and in specific contexts or use sectors will have significant impact on the actual protection offered to IoT users.

5 Conclusion

IoT devices[143] collect vast amounts of data and offer greater insight into the lives of users. To link these insights and services together, the IoT requires unique user and device identities to deliver

[139] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (2017) 17/EN WP 251 <http://www.hldataprotection.com/files/2017/10/20171013_wp251_enpdf.pdf> accessed 22 October 2017; Article 29 Data Protection Working Party, 'Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679' (2017) 17/EN WP 248.

[140] Wachter, Mittelstadt and Russell (n 108).

[141] Art 29 Wokring Party, 'Guidelines on Transparency under Regulation 2016/679; 17/EN WP260' <https://iapp.org/media/pdf/resource_center/wp29-transparency-12-12-17.pdf> accessed 4 February 2018.

[142] Unless the data breach is unlikely to result in a risk to the rights and freedoms of natural persons.

[143] Peppet (n 58).

linked and personalised databases and services for users, or a 'seamless' user experience. A fundamental tension exists between two forces in IoT systems: user privacy on the one hand, and identification technologies used to link and personalise services on the other. The collected data and monitoring facilitates the creation of profiles that lead to inferences and predictions about users. Sensor fusion (linking datasets from different devices) provides ever more detailed insight about user's private lives, and to make assumptions and predictions of their behaviour.

Linkage between IoT objects and services can inadvertently contribute to discriminatory profiling.[144] Third parties with whom the data are shared (insurance companies, employers and police) can use the data in combination with their own datasets for purposes not intended when the data was collected, or consented to by the user. Such linkage is problematic since IoT systems are built on the idea of connectivity and data sharing, hence these risks are intrinsic to IoT systems. These problems are aggravated when complex and inscrutable algorithmic systems and machine learning are used to analyse the data. Discrimination and other harmful consequences can also result from access by unauthorised third parties. The impossibility of anonymising the collected data and the risks of cyberattacks enable adversaries to use the data and the created profiles for their harmful purposes. As these characteristics of the IoT suggest, the key challenges to increase user trust and protecting privacy in IoT systems are security, privacy, data protection, profiling, and discrimination.

Thankfully, the GDPR can help to alleviate some of these risks, since data protection and privacy are primary concerns of IoT systems. Data controllers must take all reasonable steps to protect user privacy, and in particular those required by the GDPR (e.g. privacy by design and privacy by default). However, to minimise the privacy impact of the conflicts between data protection principles and identification in the IoT, the aforementioned GDPR standards urgently require further specification and implementation into the design and deployment of IoT technologies. Transparency and awareness of the possible risks and consequences of the IoT is key for users to make an informed decision if they want to use these services. Rather than pretending the problems will never occur and that privacy will always be protected, transparency and honesty about the risks that IoT pose is essential going forward.

Acknowledgements

The author is indebted to Dr. Mariarosaria Taddeo and Dr. Brent Mittelstadt of the University of Oxford, and the "Ethics, Privacy, and Trust in IoT" workshop participants who provided invaluable feedback during preparation of the manuscript. The author would also like to thank the EPSRC for the funding provided to the PETRAS consortium which made preparation of this article possible. Finally, the author would like to thank the anonymous reviewer whose feedback improved the quality of the work greatly.

Funding

This article is a deliverable of the Privacy-Enhancing and Identification-Enabling Solutions for IoT (PEIESI) project, part of the PETRAS Internet of Things research hub. PETRAS is funded by the Engineering and Physical Sciences Research Council (EPSRC), grant agreement no. EP/N023013/1. The EPSRC played no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

[144] Wachter, Mittelstadt and Floridi (n 94).

Conflict of Interest

The author declares no potential conflicts of interests.